VARIANCE COMPONENTS ESTIMATION

Skeletal Notes for A Short Course

by

S. R. Searle
Biometrics Unit
Cornell University, Ithaca, New York

Paper No. BU-719-M in the Biometrics Unit.

Table of Contents

Page

1. INTRODUCTION and the 1-WAY CLASSIFICATION - - - - - - - 1

1.1. Fixed and Random Effects - 1

1.2. History and Uses - 1

1.3. Expected Mean Squares - 1

1.4. Estimation - 2

1.5. Unbiasedness - 2

1.6. Negative Estimates - 2

1.7. Distributions, Under Normality - 2

1.8. Tests of Hypotheses - 3

1.9. Confidence Intervals - 3

1.10. Sampling Variances - 3

1.11. Unbalanced Data - 3

Expected mean squares - 3

1.12. F-statistics - 4

1.13. Maximum Likelihood - 4

2. BALANCED DATA - - - - - - - - - - - - - - - - - 5

2.1. The 2-way Crossed Classification - 5

2.2. ANOVA Rules for Balanced Data - 5

2.3. Estimation - 5

2.4. Unbiasedness - 5

2.5. Variances - 5

Unbiased estimation of variances - - - - - - - - - - - - - - - - - - 6

Minimum variance - 6

Page

3. UNBALANCED DATA: INTRODUCTION - - - - - - - - - - - 7

3.1. Estimation - 7

3.2. General Models - 7

Notations - 8

3.3. Quadratic Forms - 9

3.4. Expected Quadratic Forms - 10

Fixed effects models - 10

Mixed models - 10

Random models - 10

4. HENDERSON'S THREE METHODS - - - - - - - - - - - - 12

4.1. Henderson's Method 1 - 12

4.2. Henderson's Method 2 - 12

Estimating fixed effects - - - - - - - - - - - - - - - - - - 12

No interactions between fixed and random effects - - - - - - - - 14

Example - 14

Proof - 15

Computing procedure - 16

Proof of conditions - 19

Proof of invariance - 19

Importance - 19

4.3. Henderson's Method 3 - 20

4.4. Collected Results - 20

5. PREDICTION and MIXED MODEL EQUATIONS - - - - - - - - 21

5.1. Prediction - 21

Mood's exercise - 21

General formulation - 22

Page

5.2. The Mixed Model Equations - 23

Singularity of $\underset{\sim}{D}$ - 24

Usable solutions - 24

Generalized inverses for general solutions - - - - - - - - - - - - 24

Specific and general solutions - - - - - - - - - - - - - - - - - - - 25

Uses in estimating variance components - - - - - - - - - - - - - - - 26

6. MAXIMUM LIKELIHOOD (ML) - - - - - - - - - - - - 27

6.1. Model, Likelihood and Matrices - 27

6.2. Maximum Likelihood Equations - 27

6.3. A Computational Note - 29

6.4. Ratios of Variance Components - 29

6.5. Using the Mixed Model Equations - 30

6.6. ML Estimators - 30

6.7. Information Matrices - 31

6.8. Computing Algorithms - 32

7. RESTRICTED MAXIMUM LIKELIHOOD (REML) - - - - - - - - - 33

7.1. Error Contrasts and Their Likelihood - - - - - - - - - - - - - - - - - - 33

Error contrast - 33

Likelihood - 33

7.2. The REML Equations - 34

Direct derivation from ML - 35

7.3. Other Topics - 36

Single equation for $\hat{\sigma}_0^2$ - 36

Equations for ratios - 36

Comparisons with ML - 36

Using mixed model equations - 36

Information matrices - 36

Page

Computing algorithms - 36

Corbeil and Searle's REML - 36

8. MINQUE and MIVQUE - - - - - - - - - - - - - - - 37

8.1. MINQUE - 37

Quadratic estimation - 37

Invariance to fixed effects - 37

Unbiasedness - 37

Minimum norm - 37

Estimation equations - 38

8.2. MINQUE0 - 39

8.3. I-MINQUE - 39

8.4. MIVQUE - 40

9. DISPERSION MEAN MODEL - - - - - - - - - - - - - 41

9.1. The Model - 41

9.2. Estimation from the Model - 41

9.3. Derivation of the Model - 41

9.4. A Modified Model - 42

Topic 1

INTRODUCTION and the 1-WAY CLASSIFICATION

1.1. Fixed and Random Effects

LM 376-382 (i.e., Linear Models, pages 376-382).

1.2. History and Uses

BU-651-M, 1-8.

1.3. Expected Mean Squares

LM 385-388.

Model: $y_{ij} = \mu + \alpha_i + e_{ij}$ for $i = 1, \cdots, a$ and $j = 1, \cdots, n$

with $\alpha_i \sim (0, \sigma^2_\alpha)$, $e_{ij} \sim (0, \sigma^2_e)$ and all covariances zero.

Means: $\bar{y}_{i.} = \mu + \alpha_i + \bar{e}_{i.}$ and $\bar{y}_{..} = \mu + \bar{\alpha}_. + \bar{e}_{..}$.

Expectations:

$$SSA = \sum_{i=1}^{a} n_i(\bar{y}_{i.} - \bar{y}_{..})^2 .$$

$$E(SSA) = E\sum_{i=1}^{a} n(\alpha_i - \bar{\alpha}_. + \bar{e}_{i.} - \bar{e}_{..})^2$$

$$= n\sum_{i=1}^{a} E(\alpha_i - \bar{\alpha}_.)^2 + n\sum_{i=1}^{a} E(\bar{e}_{i.} - \bar{e}_{..})^2 + 2n\sum_{1}^{a} E(\alpha_i - \bar{\alpha}_.)(\bar{e}_{i.} - \bar{e}_{..})$$

$$= na\, E(\alpha_i^2 + \bar{\alpha}^2_{i.} - 2\alpha_i\bar{\alpha}_.) + na\, E(\bar{e}^2_{i.} + \bar{e}^2_{..} - 2\bar{e}_{i.}\bar{e}_{..}) + 0$$

$$= na\left(\sigma^2_\alpha + \frac{\sigma^2_\alpha}{a} - \frac{2\sigma^2_\alpha}{a}\right) + na\left(\frac{\sigma^2_e}{n} + \frac{\sigma^2_e}{an} - \frac{2\sigma^2_e n}{nan}\right)$$

$$= n(a - 1)\sigma^2_\alpha + (a - 1)\sigma^2_e$$

$$= (a - 1)(n\sigma^2_\alpha + \sigma^2_e) .$$

And

$$SSE = \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \bar{y}_{i.}) .$$

$$E(SSE) = E \sum_{i=1}^{a} \sum_{j=1}^{b} (e_{ij} - \bar{e}_{i.})^2$$

$$= \sum_{i=1}^{a} \sum_{j=1}^{b} E(e_{ij}^2 + \bar{e}_{i.}^2 - 2e_{ij}\bar{e}_{i.})$$

$$= \sum_{i=1}^{a} \sum_{j=1}^{b} \left(\sigma_e^2 + \frac{\sigma_e^2}{n} - \frac{2\sigma_e^2}{n}\right)$$

$$= an\sigma_e^2\left(1 - \frac{1}{n}\right)$$

$$= a(n - 1)\sigma_e^2 .$$

$$E(MSA) = E\ SSA/(a - 1) = n\sigma_\alpha^2 + \sigma_e^2 .$$

$$E(MSE) = E\ SSE/a(n - 1) = \sigma_e^2 .$$

1.4. Estimation: LM 388 (MSR_m there ≡ MSA here).

$$\left.\begin{aligned} MSA &= n\hat{\sigma}_\alpha^2 + \hat{\sigma}_e^2 \\ MSE &= \hat{\sigma}_e^2 \end{aligned}\right\} \quad \begin{aligned} \hat{\sigma}_\alpha^2 &= (MSA - MSE)/n \\ \hat{\sigma}_e^2 &= MSE . \end{aligned}$$

1.5. Unbiasedness

$$E(\hat{\sigma}_\alpha^2) = E(MSA - MSE)/n = (n\sigma_\alpha^2 + \sigma_e^2 - \sigma_e^2)/n = \sigma_\alpha^2 .$$

$$E(\hat{\sigma}_e^2) = E(MSE) = \sigma_e^2 .$$

1.6. Negative Estimates

LM 406-408.

1.7. Distributions, Under Normality

LM 410.

1.8. Tests of Hypotheses

$$F = \frac{MSA}{n\sigma_\alpha^2 + \sigma_e^2} \bigg/ \frac{MSE}{\sigma_e^2} \sim F_{a-1, a(n-1)} .$$

Under $H : \sigma_\alpha^2 = 0, \qquad F = \frac{MSA}{MSE}$.

1.9. Confidence Intervals

LM 414, Table 9.14.

1.10. Sampling Variances

LM 416.

1.11. Unbalanced Data

Unbalanced ≡ unequal numbers of observations in the classes.

Model: $y_{ij} = \mu + \alpha_i + e_{ij}$ for $i = 1, \cdots, a$ and $j = 1, \cdots, n_i$.

Expected mean squares

$$SSA = \sum_{i=1}^{a} n_i(\bar{y}_{i.} - \bar{y}_{..})^2 .$$

$$E(SSA) = E\sum_{i=1}^{a} n_i(\alpha_i - \bar{\alpha}_{i.} + \bar{e}_{i.} - \bar{e}_{..})$$

$$= E\sum_{i=1}^{a} n_i\left(\alpha_i - \frac{\Sigma n_i\alpha_i}{n_.} + \bar{e}_{i.} - \bar{e}_{..}\right)^2$$

$$= \sum_{i=1}^{a} n_i E\left(\alpha_i - \frac{\Sigma n_i\alpha_i}{n_.}\right)^2 + \sum_{i=1}^{a} n_i E(\bar{e}_{i.} - \bar{e}_{..})^2$$

$$+ \sum_{i=1}^{a} n_i E\left(\alpha_i - \frac{\Sigma n_i\alpha_i}{n_.}\right)(\bar{e}_{i.} - \bar{e}_{..})$$

$$= \sum_{i=1}^{a} n_i E\left[\alpha_i^2 + \frac{(\Sigma n_i\alpha_i)^2}{n_.^2} - 2\,\frac{\alpha_i(\Sigma n_i\alpha_i)}{n_.}\right] + \sum_{i=1}^{a} n_i E(\bar{e}_{i.}^2 + \bar{e}_{..}^2 - 2\bar{e}_{i.}\bar{e}_{..}) + 0$$

$$= \sum_{i=1}^{a} n_i\left(\sigma_\alpha^2 + \frac{\Sigma n_i^2\sigma_\alpha^2}{n_.^2} - \frac{2n_i\sigma_\alpha^2}{n_.}\right) + \sum_{i=1}^{a} n_i E\left(\frac{\sigma_e^2}{n_i} + \frac{\sigma_e^2}{n_.} - \frac{2n_i\sigma_e^2}{n_i n_.}\right)$$

$$= \sigma_\alpha^2\left(n_. + \frac{\Sigma n_i^2}{n_.} - \frac{2\Sigma n_i^2}{n_.}\right) + \sigma_e^2(a + 1 - 2)$$

$$= (n_. - \Sigma n_i^2/n_.)\sigma_\alpha^2 + (a - 1)\sigma_e^2 .$$

$$\text{SSE} = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 .$$

$$E(\text{SSE}) = E \sum_{i=1}^{a} \sum_{j=1}^{n_i} (e_{ij} - \bar{e}_{i.})^2$$

$$= \sum_{i=1}^{a} \sum_{j=1}^{n_i} E(e_{ij}^2 + \bar{e}_{i.}^2 - 2e_{ij}\bar{e}_{i.})$$

$$= \sum_{i=1}^{a} \sum_{j=1}^{n_i} \left(\sigma_e^2 + \frac{\sigma_e^2}{n_i} - \frac{2\sigma_e^2}{n_i}\right)$$

$$= \sum_{i=1}^{a} \sigma_e^2(n_i + 1 - 2)$$

$$= \sigma_e^2(n_. - a) .$$

$$E(\text{MSA}) = E\ \text{SSA}/(a - 1) = \frac{n_. - \Sigma n_i^2/n_.}{a - 1}\sigma_\alpha^2 + \sigma_e^2 .$$

$$E(\text{MSE}) = E\ \text{SSA}/(n_. - a) = \sigma_e^2 .$$

1.12. F-statistics

Utility of F = MSA/MSE in the 1-way classification
(under the usual normality assumptions)

Model	Data	
	Balanced	Unbalanced
Fixed	F tests $H: \alpha_i$'s all equal	F tests $H: \alpha_i$'s all equal
Random	F tests $H: \sigma_\alpha^2 = 0$	F is not distributed as an F-variable*

* SSA is not distributed (proportionally) as a χ^2-variable, but is a weighted sum of one-degree-of-freedom χ^2's, with the weights being functions of σ_α^2, σ_e^2 and the n_i's.

1.13. Maximum Likelihood

LM 418-419.

Topic 2

BALANCED DATA

2.1. The 2-way Crossed Classification

Model: $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$

for $i = 1, \cdots, a$, $j = 1, \cdots, b$, and $k = 1, \cdots, n$.

This is for a rows, b columns and n observations per cell.

LM 395-404 (transparencies).

The mixed model discussion of LM 400-404 is also dealt with in Hocking (*American Statistician*, 27, 148, 1973). Equivalences are as follows:

LM	Hocking
Table 9.9	Model II
Table 9.10	Models I and III

2.2. ANOVA Rules for Balanced Data

LM 389-394 (transparencies).

2.3. Estimation   LM 405-406

Definitions: $\underset{\sim}{m}$ = vector of mean squares

$\underset{\sim}{\sigma}^2$ = vector of variance components

$E(\underset{\sim}{m}) = \underset{\sim}{P}\underset{\sim}{\sigma}^2$, defines $\underset{\sim}{P}$.

Estimators: $\hat{\underset{\sim}{\sigma}}^2 = \underset{\sim}{P}^{-1}\underset{\sim}{m}$.

2.4. Unbiasedness

$$E(\hat{\underset{\sim}{\sigma}}^2) = E(\underset{\sim}{P}^{-1}\underset{\sim}{m}) = \underset{\sim}{P}^{-1}E(\underset{\sim}{m}) = \underset{\sim}{P}^{-1}\underset{\sim}{P}\underset{\sim}{\sigma}^2 = \underset{\sim}{\sigma}^2 .$$

2.5. Variances   LM 416-417

$$\operatorname{var}(\hat{\underset{\sim}{\sigma}}^2) = \underset{\sim}{P}^{-1}\operatorname{var}(\underset{\sim}{m})\underset{\sim}{P}^{-1'} .$$

$\hat{\sigma}^2$'s are linear combinations of sums of squares.

Under normality, distributions of sums of squares are proportional to χ^2's.

In ANOVA of balanced data, sums of squares are uncorrelated and

$$v(MS) = \frac{2[E(MS)]^2}{d.f.} .$$

Hence

$$\text{var}(\underset{\sim}{m}) \equiv \underset{\sim}{D} = \text{diag}\left\{\frac{2[E(M_i)]^2}{f_i}\right\} \qquad (69),\ \text{LM } 417$$

and

$$\text{var}(\underset{\sim}{\hat{\sigma}^2}) = \underset{\sim}{P}^{-1}\underset{\sim}{D}\underset{\sim}{P}^{-1'} . \qquad (70),\ \text{LM } 417$$

Note that although each $\hat{\sigma}^2$ is a linear combination of mean squares, they do not all occur with positive coefficients, and so the $\hat{\sigma}^2$'s are not distributed as sums of χ^2-variables.

Unbiased estimation of variances

$$\underset{\sim}{D}_1 = \text{diag}\left\{\frac{2M_i^2}{f_i}\right\}$$

$$\widehat{\text{var}}(\underset{\sim}{\hat{\sigma}^2}) = \underset{\sim}{P}^{-1}\underset{\sim}{D}_1\underset{\sim}{P}^{-1'} \quad \text{is not unbiased.} \qquad (72),\ \text{LM } 417$$

$$\underset{\sim}{D}_2 = \text{diag}\left\{\frac{2M_i^2}{f_i + 2}\right\}$$

$$\widehat{\text{var}}(\underset{\sim}{\hat{\sigma}^2}) = \underset{\sim}{P}^{-1}\underset{\sim}{D}_2\underset{\sim}{P}^{-1'} \quad \text{is unbiased.} \qquad (74),\ \text{LM } 417$$

See bottom of LM 417 for reasons.

Minimum variance: LM 406

BQUE: Graybill and Hultquist (Ann. Math. Stat., 1961).

BUE, under normality: Graybill (Ann. Math. Stat., 1954).

BUE, under arbitrary kurtosis: Anderson et al. (BU-691-M, 1980).

Topic 3

UNBALANCED DATA: INTRODUCTION

3.1. Estimation

Balanced data: use $\underset{\sim}{m}$ = vector of mean squares.

Unbalanced data: use $\underset{\sim}{q}$ = vector of quadratic forms.

$$\underset{\sim}{q} = \{q_i\} = \{\underset{\sim}{y}'\underset{\sim}{A}_i\underset{\sim}{y}\}$$

$$E\,\underset{\sim}{q} = \underset{\sim}{C}\sigma^2$$

$$\underset{\sim}{\hat{\sigma}^2} = \underset{\sim}{C}^{-1}\underset{\sim}{q}\ .$$

Question: What to use for q_i's?

Answer: "mean squares".

Questions: What mean squares?

Why mean squares?

Further questions: Properties of $\underset{\sim}{\hat{\sigma}^2}$?

Are some mean squares better than others?

3.2. General Models

Fixed effects: $\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{b} + \underset{\sim}{e}$

Mixed model: $\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{b} + \underset{\sim}{Z}_1\underset{\sim}{u}_1 + \underset{\sim}{Z}_2\underset{\sim}{u}_2 + \cdots + \underset{\sim}{Z}_c\underset{\sim}{u}_c + \underset{\sim}{e}$

$\underset{\sim}{b}$: fixed effects

$\underset{\sim}{u}_i$: effects for levels of random factor i, be it a main effects factor, a nested factor or an interaction factor.

Notations:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \cdots + \mathbf{Z}_c\mathbf{u}_c + \mathbf{e}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \cdots + \mathbf{Z}_c\mathbf{u}_c + \mathbf{e}$$

(e.g., Hartley and Rao, Henderson)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}_1\mathbf{b}_1 + \mathbf{Z}_2\mathbf{b}_2 + \cdots + \mathbf{Z}_c\mathbf{b}_c + \mathbf{e}$$

(e.g., Harville, Searle)

$$\mathbf{y} = \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_A\mathbf{b}_A + \mathbf{X}_B\mathbf{b}_B + \cdots + \mathbf{X}_K\mathbf{b}_K + \mathbf{e} \qquad \text{(LM 423).}$$

For consistency with BU-673-M (hereinafter referenced as NVC), we will use

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}_1\mathbf{b}_1 + \mathbf{Z}_2\mathbf{b}_2 + \cdots + \mathbf{Z}_c\mathbf{b}_c + \mathbf{e}$$

$$= \mathbf{X}\boldsymbol{\alpha} + \sum_{i=1}^{c} \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}$$

$$= \mathbf{X}\boldsymbol{\alpha} + [\mathbf{Z}_1 \;\; \mathbf{Z}_2 \; \cdots \; \mathbf{Z}_c] \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_c \end{bmatrix} + \mathbf{e}$$

$$= \mathbf{X}\boldsymbol{\alpha} + \mathbf{Zb} + \mathbf{e}$$

$$\text{with } \mathbf{Z} = [\mathbf{Z}_1 \;\; \mathbf{Z}_2 \; \cdots \; \mathbf{Z}_c] \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_c \end{bmatrix}.$$

Also, on defining

$$\mathbf{b}_0 \equiv \mathbf{e} \quad \text{and} \quad \mathbf{Z}_0 = \mathbf{I}_N ,$$

$$\underset{\sim}{y} = \underset{\sim}{X\alpha} + \underset{\sim}{Z}_1\underset{\sim}{b}_1 + \underset{\sim}{Z}_2\underset{\sim}{b}_2 + \cdots + \underset{\sim}{Z}_c\underset{\sim}{b}_c + \underset{\sim}{Z}_0\underset{\sim}{b}_0$$

$$= \underset{\sim}{X\alpha} + \underset{\sim}{Z}_0\underset{\sim}{b}_0 + \underset{\sim}{Z}_1\underset{\sim}{b}_1 + \cdots \quad \cdots + \underset{\sim}{Z}_c\underset{\sim}{b}_c$$

$$= \underset{\sim}{X\alpha} + \sum_{i=0}^{c} \underset{\sim}{Z}_i\underset{\sim}{b}_i$$

$$= \underset{\sim}{X\alpha} + [\underset{\sim}{Z}_0 \quad \underset{\sim}{Z}_1 \quad \cdots \quad \underset{\sim}{Z}_c] \begin{bmatrix} \underset{\sim}{b}_0 \\ \underset{\sim}{b}_1 \\ \vdots \\ \underset{\sim}{b}_c \end{bmatrix}$$

$$= \underset{\sim}{X\alpha} + \underset{\sim}{\dot{Z}\dot{b}}$$

$$\text{with } \underset{\sim}{\dot{Z}} = [\underset{\sim}{Z}_0 \quad \underset{\sim}{Z}] \quad \text{and} \quad \underset{\sim}{\dot{b}} = \begin{bmatrix} \underset{\sim}{b}_0 \\ \underset{\sim}{b} \end{bmatrix} .$$

The dot notation is used in NVC. It is not conventional. Some authors use $\underset{\sim}{Z}$ as defined here and some use it for what is here called $\underset{\sim}{\dot{Z}}$.

3.3. Quadratic Forms

$\underset{\sim}{y}$ = vector of data, of order N

$\underset{\sim}{A}$ = symmetric matrix

$E(\underset{\sim}{y}) = \underset{\sim}{\mu} \qquad \text{var}(\underset{\sim}{y}) = \underset{\sim}{V}$.

Mean: $E(\underset{\sim}{y}'\underset{\sim}{Ay}) = \text{tr}(\underset{\sim}{AV}) + \underset{\sim}{\mu}'\underset{\sim}{A\mu}$ (LM 55)

Variance: under normality

$$v(\underset{\sim}{y}'\underset{\sim}{Ay}) = 2\text{tr}(\underset{\sim}{AV})^2 + 4\underset{\sim}{\mu}'\underset{\sim}{AVA\mu} \qquad \text{(LM 57)}$$

Covariances: under normality

$$\text{cov}(\underset{\sim}{y}'\underset{\sim}{Ay},\ \underset{\sim}{y}'\underset{\sim}{By}) = 2\text{tr}(\underset{\sim}{AVBV}) + 4\underset{\sim}{\mu}'\underset{\sim}{AVB\mu} . \qquad \text{(LM 66)}$$

3.4. Expected Quadratic Forms

Fixed effects models:

$$\underset{\sim}{y} = \underset{\sim}{Xb} + \underset{\sim}{e} \sim (\underset{\sim}{Xb},\ \sigma_e^2 \underset{\sim}{I})$$

$$E(\underset{\sim}{y}'\underset{\sim}{Ay}) = \underset{\sim}{b}'\underset{\sim}{X}'\underset{\sim}{AXb} + \sigma_e^2 \operatorname{tr}(\underset{\sim}{A}) \qquad \text{(LM 422)}$$

Mixed models:

$$\underset{\sim}{y} = \underset{\sim}{X\alpha} + \underset{\sim}{Zb} + \underset{\sim}{e} \sim (\underset{\sim}{X\alpha},\ \underset{\sim}{V})$$

where

$$\underset{\sim}{V} = \operatorname{var}(\underset{\sim}{y}) = \operatorname{var}(\underset{\sim}{Zb} + \underset{\sim}{e}) \ .$$

NVC 6-9:

$$\underset{\sim}{D} = \bigoplus_{i=1}^{c} \sigma_i^2 \underset{\sim}{I}_{q_i}$$

$$= \operatorname{diag}\{\sigma_1^2 \underset{\sim}{I}_{q_1} \cdots \sigma_c^2 \underset{\sim}{I}_{q_c}\}$$

$$\underset{\sim}{V} = \underset{\sim}{ZDZ}' + \sigma_e^2 \underset{\sim}{I}_N$$

$$= \sum_{i=1}^{c} \sigma_i^2 \underset{\sim}{Z}_i \underset{\sim}{Z}_i' + \sigma_e^2 \underset{\sim}{I}_N \qquad \begin{cases} \text{NVC 8, (1.18)} \\ \text{LM 423, (9)} \end{cases}$$

$$E(\underset{\sim}{y}'\underset{\sim}{Ay}) = \underset{\sim}{\alpha}'\underset{\sim}{X}'\underset{\sim}{AX\alpha} + \operatorname{tr}\left[\underset{\sim}{A}\left(\sum_{i=1}^{c} \sigma_i^2 \underset{\sim}{Z}_i \underset{\sim}{Z}_i' + \sigma_e^2 \underset{\sim}{I}\right)\right]$$

$$= \underset{\sim}{\alpha}'\underset{\sim}{X}'\underset{\sim}{AX\alpha} + \sum_{i=1}^{c} \sigma_i^2 \operatorname{tr}(\underset{\sim}{AZ}_i \underset{\sim}{Z}_i') + \sigma_e^2 \operatorname{tr}(\underset{\sim}{A}) \ . \qquad \text{[LM 424, (10)]}$$

Random models: $\underset{\sim}{X\alpha} = \underset{\sim}{\mathbf{1}}\mu$

$$E(\underset{\sim}{y}'\underset{\sim}{Ay}) = \mu^2 \underset{\sim}{\mathbf{1}}'\underset{\sim}{A\mathbf{1}} + \sum_{i=1}^{c} \sigma_i^2 \operatorname{tr}(\underset{\sim}{AZ}_i \underset{\sim}{Z}_i') + \sigma_e^2 \operatorname{tr}(\underset{\sim}{A}) \ . \qquad \text{[LM 424, (11)]}$$

Quadratic forms $\underset{\sim}{y}'\underset{\sim}{Ay}$ used in methods of estimating variance components that are based on ANOVA mean squares (or something akin thereto), are often such that

$\underset{\sim}{1}'\underset{\sim}{A}\underset{\sim}{1} = \underset{\sim}{0}$ and so the term in μ^2 drops out of $E(\underset{\sim}{y}'\underset{\sim}{A}\underset{\sim}{y})$. More generally, other forms of estimation use an $\underset{\sim}{A}$ such that $\underset{\sim}{A}\underset{\sim}{X} = \underset{\sim}{0}$ and so the term $\underset{\sim}{\alpha}'\underset{\sim}{X}'\underset{\sim}{A}\underset{\sim}{X}\underset{\sim}{\alpha}$ drops out of $E(\underset{\sim}{y}'\underset{\sim}{A}\underset{\sim}{y})$ for mixed models.

Topic 4

HENDERSON'S THREE METHODS

Henderson (Biometrics, 1953) gave three methods of estimating variance components from unbalanced data. These methods are based upon analysis of variance concepts. Searle (Biometrics, 1968) reformulated the methods in matrix notation, generalized Method 2, and suggested it had no unique usage for any given set of data. But Henderson, Searle and Schaeffer (Biometrics, 1974) show that this suggestion is wrong and they also give simplified calculation procedures.

4.1. Henderson's Method 1

LM 425-440 (transparencies). Comment on computability: always feasible.

4.2. Henderson's Method 2

Model: $\underset{\sim}{y} = \mu\underset{\sim}{1} + \underbrace{\underset{\sim}{X}_f\underset{\sim}{b}_f}_{\substack{\uparrow \\ \text{fixed} \\ \text{effects}}} + \underbrace{\underset{\sim}{X}_r\underset{\sim}{b}_r}_{\substack{\uparrow \\ \text{random} \\ \text{effects}}} + \underset{\sim}{e}$

Estimate $\underset{\sim}{b}_f$: $\tilde{\underset{\sim}{b}}_f = \underset{\sim}{L}\underset{\sim}{y}$ for some $\underset{\sim}{L}$.

Adjust data for $\tilde{\underset{\sim}{b}}_f$:

$$\begin{aligned}\underset{\sim}{z} &= \underset{\sim}{y} - \underset{\sim}{X}_f\tilde{\underset{\sim}{b}}_f \\ &= (\underset{\sim}{I} - \underset{\sim}{X}_f\underset{\sim}{L})\underset{\sim}{y} \ .\end{aligned}$$

Estimating fixed effects

Suppose the model for $\underset{\sim}{z}$ were

$$\underset{\sim}{z} = \mu^*\underset{\sim}{1} + \underset{\sim}{X}_r\underset{\sim}{b}_r + \underset{\sim}{Z}\underset{\sim}{e}$$

for some scalar μ^* and some matrix $\underset{\sim}{Z}$. Then, apart from $\underset{\sim}{Z}\underset{\sim}{e}$, estimation of variance components could be made from $\underset{\sim}{z}$ with the random effects occurring in $\underset{\sim}{z}$ just as they do in $\underset{\sim}{y}$, but with no fixed effects in $\underset{\sim}{z}$. In other words, Method 1 could be applied to $\underset{\sim}{z}$ and coefficients of σ^2's in $E(\underset{\sim}{q})$ would, apart from those of σ_e^2, be exactly the same as in $E(\underset{\sim}{q})$ of $\underset{\sim}{y}$. This is the objective of Method 2.

Model for $\underset{\sim}{z}$:

$$\underset{\sim}{z} = \mu(\underset{\sim}{1} - \underset{\sim}{X}_f\underset{\sim}{L}\underset{\sim}{1}) + (\underset{\sim}{X}_f - \underset{\sim}{X}_f\underset{\sim}{L}\underset{\sim}{X}_f)\underset{\sim}{b}_f + (\underset{\sim}{X}_r - \underset{\sim}{X}_f\underset{\sim}{L}\underset{\sim}{X}_r)\underset{\sim}{b}_r + (\underset{\sim}{I} - \underset{\sim}{X}_f\underset{\sim}{L})\underset{\sim}{e} .$$

The most direct way to eliminate the fixed effects, $\underset{\sim}{b}_f$, from the model for $\underset{\sim}{z}$ is to choose $\underset{\sim}{L}$ so that $\underset{\sim}{X}_f - \underset{\sim}{X}_f\underset{\sim}{L}\underset{\sim}{X}_f = \underset{\sim}{0}$, i.e., $\underset{\sim}{X}_f = \underset{\sim}{X}_f\underset{\sim}{L}\underset{\sim}{X}_f$, or $\underset{\sim}{L}$ a generalized inverse of $\underset{\sim}{X}_f$. This has been called the general method 2 by Searle (1968, 1971).

But Henderson's Method 2 is based not on eliminating the fixed effects but on combining them with μ in a particular way. Suppose

(i) $\underset{\sim}{X}_f\underset{\sim}{L}$ has its row sums equal.

Then $\underset{\sim}{X}_f\underset{\sim}{L}\underset{\sim}{1} = c_1\underset{\sim}{1}$ for some scalar c_1.

(ii) $\underset{\sim}{X}_f - \underset{\sim}{X}_f\underset{\sim}{L}\underset{\sim}{X}_f$ has every row the same, $\underset{\sim}{\rho}'$ say.

Then $\underset{\sim}{X}_f - \underset{\sim}{X}_f\underset{\sim}{L}\underset{\sim}{X}_f = \underset{\sim}{1}\underset{\sim}{\rho}'$.

(iii) $\underset{\sim}{X}_f\underset{\sim}{L}\underset{\sim}{X}_r = \underset{\sim}{0}$.

Then the model for $\underset{\sim}{z}$ becomes

$$\begin{aligned}
\underset{\sim}{z} &= \mu(\underset{\sim}{1} - c_1\underset{\sim}{1}) + \underset{\sim}{1}\underset{\sim}{\rho}'\underset{\sim}{b}_f + \underset{\sim}{X}_r\underset{\sim}{b}_r + (\underset{\sim}{I} - \underset{\sim}{X}_f\underset{\sim}{L})\underset{\sim}{e} \\
&= [\mu(1 - c_1) + \underset{\sim}{\rho}'\underset{\sim}{b}_f]\underset{\sim}{1} + \underset{\sim}{X}_r\underset{\sim}{b}_r + (\underset{\sim}{I} - \underset{\sim}{X}_f\underset{\sim}{L})\underset{\sim}{e} \\
&= \mu^* + \underset{\sim}{X}_r\underset{\sim}{b}_r + \underset{\sim}{Z}\underset{\sim}{e}
\end{aligned}$$

$$\text{with} \quad \mu^* = \mu(1 - c_1) + \underset{\sim}{\rho}'\underset{\sim}{b}_f$$

$$\underset{\sim}{Z} = \underset{\sim}{I} - \underset{\sim}{X}_f\underset{\sim}{L} .$$

In this way, the fixed effects are combined with μ. Henderson's Method 2 uses an $\underset{\sim}{L}$ that satisfies (i), (ii) and (iii).

No interactions between fixed and random effects

Theorem

The conditions (i), (ii) and (iii) that $\mathbf{L}$ must satisfy imply that the model for $\mathbf{y}$ must not contain interactions between the fixed effects and random effects.

This is so, whether one chooses to define such interactions as random effects (the usual case) or as fixed effects. The proof was given for the first time in Searle (1968). It goes as follows. But first, an example to illustrate certain relationships between $\mathbf{X}_f$ and $\mathbf{X}_r$ when interactions exist.

Example: 2-way crossed classification.

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij} = \mathbf{Xb}$$

n_{ij}'s	
2	3
1	2

Model matrix:

$$\mathbf{X} = \left[\begin{array}{c|cc|cc|cccc}
1 & 1 & . & 1 & . & 1 & . & . & . \\
1 & 1 & . & 1 & . & 1 & . & . & . \\
1 & 1 & . & . & 1 & . & 1 & . & . \\
1 & 1 & . & . & 1 & . & 1 & . & . \\
1 & 1 & . & . & 1 & . & 1 & . & . \\
1 & . & 1 & 1 & . & . & . & 1 & . \\
1 & . & 1 & . & 1 & . & . & . & 1 \\
1 & . & 1 & . & 1 & . & . & . & 1
\end{array}\right] \equiv [\mathbf{1} \quad \mathbf{X}_A \quad \mathbf{X}_B \quad \mathbf{X}_{AB}]$$

$\mathbf{X}_A$ has columns that are sums of columns of $\mathbf{X}_{AB}$

$\mathbf{X}_B$ " " " " " " " " " .

Take α's as fixed and β's as random. Suppose interactions are random. Then

$$\underset{\sim}{X}_f = [\underset{\sim}{1} \quad \underset{\sim}{X}_A] \quad \text{and} \quad \underset{\sim}{X}_r = [\underset{\sim}{X}_B \quad \underset{\sim}{X}_{AB}] .$$

Therefore some columns of $\underset{\sim}{X}_f$ are sums of certain columns of $\underset{\sim}{X}_r$. (Actually the relationship between $\underset{\sim}{X}_f$ and $\underset{\sim}{X}_r$ is much more specific in this, the 2-factor model, but this statement is true quite generally, no matter how many factors.) Hence, apart from permuting columns, $\underset{\sim}{X}_f$ could be partitioned as

$$\underset{\sim}{X}_f = [\underset{\sim}{X}_{f1} \quad \underset{\sim}{X}_{f2}]$$

where $\underset{\sim}{X}_{f2}$ represents those columns that are sums of columns of $\underset{\sim}{X}_{AB}$ and so

$$\underset{\sim}{X}_{f2} = \underset{\sim}{X}_r\underset{\sim}{M} \quad \text{for some } \underset{\sim}{M}. \tag{1}$$

Similarly, if interactions are taken as fixed effects we can partition $\underset{\sim}{X}_r$ as

$$\underset{\sim}{X}_r = [\underset{\sim}{X}_{r1} \quad \underset{\sim}{X}_{r2}]$$

and have

$$\underset{\sim}{X}_{r2} = \underset{\sim}{X}_f\underset{\sim}{K} \quad \text{for some } \underset{\sim}{K}. \tag{2}$$

Proof. Suppose interactions are random. Then

$$\begin{aligned}\underset{\sim}{X}_f\underset{\sim}{L}\underset{\sim}{X}_{f2} &= \underset{\sim}{X}_f\underset{\sim}{L}\underset{\sim}{X}_r\underset{\sim}{M}, \quad \text{from (1)} \\ &= \underset{\sim}{0} \quad \text{from condition (iii).}\end{aligned}$$

Therefore in condition (ii)

$$\begin{aligned}&[\underset{\sim}{X}_{f1} \quad \underset{\sim}{X}_{f2}] - \underset{\sim}{X}_f\underset{\sim}{L}[\underset{\sim}{X}_{f1} \quad \underset{\sim}{X}_{f2}] \\ &\quad = [\underset{\sim}{X}_{f1} \quad \underset{\sim}{X}_{f2}] - [\underset{\sim}{X}_f\underset{\sim}{L}\underset{\sim}{X}_{f1} \quad \underset{\sim}{0}] \\ &\quad = [\underset{\sim}{X}_{f1} - \underset{\sim}{X}_f\underset{\sim}{L}\underset{\sim}{X}_{f1} \quad \underset{\sim}{X}_{f2}] \text{ has every row the same.}\end{aligned}$$

This means $\underset{\sim}{X}_{f2}$ has every row the same — which is meaningless. Therefore Henderson's Method 2 cannot have interactions between fixed and random effects if those interactions are treated as random.

Now suppose the interactions are fixed. Then in condition (iii)

$$\underset{\sim}{X}_f \underset{\sim}{L} \underset{\sim}{X}_r = \underset{\sim}{X}_f \underset{\sim}{L} [\underset{\sim}{X}_{r1} \quad \underset{\sim}{X}_{r2}] = \underset{\sim}{0} \Rightarrow \underset{\sim}{X}_f \underset{\sim}{L} \underset{\sim}{X}_{r2} = \underset{\sim}{0} \ .$$

Hence, using (2),

$$\underset{\sim}{X}_f \underset{\sim}{L} \underset{\sim}{X}_{r2} = \underset{\sim}{X}_f \underset{\sim}{L} \underset{\sim}{X}_f \underset{\sim}{K} = \underset{\sim}{0} \ .$$

Therefore, on post-multiplying (ii) by $\underset{\sim}{K}$,

$$\underset{\sim}{X}_f \underset{\sim}{K} - \underset{\sim}{X}_f \underset{\sim}{L} \underset{\sim}{X}_f \underset{\sim}{K} = \underset{\sim}{\mathbf{1}} \underset{\sim}{\rho}' \underset{\sim}{K} \ ,$$

i.e.,

$$\underset{\sim}{X}_f \underset{\sim}{K} = \underset{\sim}{\mathbf{1}} (\underset{\sim}{\rho}' \underset{\sim}{K})$$

or, by (2),

$$\underset{\sim}{X}_{r2} = \underset{\sim}{\mathbf{1}} (\underset{\sim}{\rho}' \underset{\sim}{K}) \ .$$

This means every row of $\underset{\sim}{X}_{r2}$ is the same — again, a meaningless conclusion. Q.E.D.

The consequence of this theorem is that be they treated as fixed or random, interactions can be part of the model when Henderson's Method 2 is used only if they are interactions of fixed effects with each other, or of random effects with each other, and not of fixed effects with random effects.

Computing procedure

The procedure is as follows.

(a) Use the model

$$E(\underset{\sim}{y}) = \mu \underset{\sim}{\mathbf{1}} + \underset{\sim}{X}_f \underset{\sim}{\beta}_f + \underset{\sim}{X}_r \underset{\sim}{\beta}_r$$

as if $\underset{\sim}{\beta}_r$ were fixed effects. The normal equations are

$$\begin{bmatrix} \underset{\sim}{1}'\underset{\sim}{1} & \underset{\sim}{1}'\underset{\sim}{X}_f & \underset{\sim}{1}'\underset{\sim}{X}_r \\ \underset{\sim}{X}_f'\underset{\sim}{1} & \underset{\sim}{X}_f'\underset{\sim}{X}_f & \underset{\sim}{X}_f'\underset{\sim}{X}_r \\ \underset{\sim}{X}_r'\underset{\sim}{1} & \underset{\sim}{X}_r'\underset{\sim}{X}_f & \underset{\sim}{X}_r'\underset{\sim}{X}_r \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\underset{\sim}{b}}_f \\ \hat{\underset{\sim}{b}}_r \end{bmatrix} = \begin{bmatrix} \underset{\sim}{1}'\underset{\sim}{y} \\ \underset{\sim}{X}_f'\underset{\sim}{y} \\ \underset{\sim}{X}_r'\underset{\sim}{y} \end{bmatrix} .$$

(b) Take $\hat{\mu} \equiv 0$. This reduces the equation to

$$\begin{bmatrix} \underset{\sim}{X}_f'\underset{\sim}{X}_f & \underset{\sim}{X}_f'\underset{\sim}{X}_r \\ \underset{\sim}{X}_r'\underset{\sim}{X}_f & \underset{\sim}{X}_r'\underset{\sim}{X}_r \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{b}}_f \\ \hat{\underset{\sim}{b}}_r \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}_f'\underset{\sim}{y} \\ \underset{\sim}{X}_r'\underset{\sim}{y} \end{bmatrix} .$$

These equations are not of full rank, and have many solutions, obtainable by using generalized inverses of

$$\underset{\sim}{X}'\underset{\sim}{X} = \begin{bmatrix} \underset{\sim}{X}_f'\underset{\sim}{X}_f & \underset{\sim}{X}_f'\underset{\sim}{X}_r \\ \underset{\sim}{X}_r'\underset{\sim}{X}_f & \underset{\sim}{X}_r'\underset{\sim}{X}_r \end{bmatrix} .$$

(c) For Henderson's Method 2, a generalized inverse of $\underset{\sim}{X}'\underset{\sim}{X}$ is chosen as follows (Searle, 1968, pp. 758-760):

Strike out from $\underset{\sim}{X}'\underset{\sim}{X}$ as many rows and columns as is necessary to leave a matrix of full rank (equal to the rank of $\underset{\sim}{X}'\underset{\sim}{X}$). In striking out rows and columns, as many as possible <u>must</u> be rows and columns through $\underset{\sim}{X}_f'\underset{\sim}{X}_f$. (This is the <u>crux</u> of Henderson's Method 2.) Call the remaining full rank sub-matrix $\underset{\sim}{D}$. Within $\underset{\sim}{X}'\underset{\sim}{X}$ replace $\underset{\sim}{D}$ by $\underset{\sim}{D}^{-1}$, element for element, and in the struck out rows and columns put zeros. The result is a generalized inverse

$$(\underset{\sim}{X}'\underset{\sim}{X})^- = \begin{bmatrix} \underset{\sim}{P}_{11} & \underset{\sim}{P}_{12} \\ \underset{\sim}{P}_{21} & \underset{\sim}{P}_{22} \end{bmatrix}$$

giving

$$\tilde{\underset{\sim}{b}}_f = \underset{\approx}{L}\underset{\sim}{y} = [\underset{\sim}{P}_{11} \quad \underset{\sim}{P}_{12}]\underset{\sim}{X}'\underset{\sim}{y} .$$

(d) Carry out Henderson's Method 1 on

$$\underset{\sim}{z} = \underset{\sim}{y} - \underset{\sim}{X}_f \tilde{\underset{\sim}{b}}_f = \mu^* \underset{\sim}{1} + \underset{\sim}{X}_r \underset{\sim}{b}_r + (\underset{\sim}{I} - \underset{\sim}{X}_f \underset{\sim}{L}) \underset{\sim}{e} \ ,$$

using $\underset{\sim}{z}'\underset{\sim}{A}\underset{\sim}{z}$ for each $\underset{\sim}{A}$ that would be used in Method 1 estimation from $\underset{\sim}{y}$ if there were no fixed effects. $E(\underset{\sim}{z}'\underset{\sim}{A}\underset{\sim}{z})$ will contain the same terms in the variance components as does $E(\underset{\sim}{y}'\underset{\sim}{A}\underset{\sim}{y})$, except for terms in σ_e^2.

(e) Terms in σ_e^2 are calculated as follows (Henderson et al., 1974, pp. 586-588):

Suppose the term in σ_e^2 in $E(\underset{\sim}{y}'\underset{\sim}{A}\underset{\sim}{y})$ is $k_A \sigma_e^2$. Then the term in σ_e^2 in $E(\underset{\sim}{z}'\underset{\sim}{A}\underset{\sim}{z})$ is $(k_A + \delta_A)\sigma_e^2$.

δ_A is defined as follows.

Partitioning: $\underset{\sim}{X}_f = [\underset{\sim}{F}_1 \;\; \underset{\sim}{F}_2] \qquad \underset{\sim}{X}_r = [\underset{\sim}{R}_1 \;\; \underset{\sim}{R}_2]$

$$\underset{\sim}{X}'\underset{\sim}{X} = \begin{bmatrix} \underset{\sim}{F}_1'\underset{\sim}{F}_1 & \underset{\sim}{F}_1'\underset{\sim}{F}_2 & \underset{\sim}{F}_1'\underset{\sim}{R}_1 & \underset{\sim}{F}_1'\underset{\sim}{R}_2 \\ & \underset{\sim}{F}_2'\underset{\sim}{F}_2 & \underset{\sim}{F}_2'\underset{\sim}{R}_1 & \underset{\sim}{F}_2'\underset{\sim}{R}_2 \\ & & \underset{\sim}{R}_1'\underset{\sim}{R}_1 & \underset{\sim}{R}_1'\underset{\sim}{R}_2 \\ \text{sym} & & & \underset{\sim}{R}_2'\underset{\sim}{R}_2 \end{bmatrix} .$$

The partitioning is defined such that (apart from permuting rows and columns) the rows and columns of $\underset{\sim}{X}'\underset{\sim}{X}$ that are struck out to obtain the required form of $(\underset{\sim}{X}'\underset{\sim}{X})^-$ are those through $\underset{\sim}{F}_1'\underset{\sim}{F}_1$ and $\underset{\sim}{R}_2'\underset{\sim}{R}_2$. Hence (Searle, 1968, p. 758 and 775)

$$(\underset{\sim}{X}'\underset{\sim}{X})^- = \begin{bmatrix} \underset{\sim}{P}_{11} & \underset{\sim}{P}_{12} \\ \underset{\sim}{P}_{21} & \underset{\sim}{P}_{22} \end{bmatrix} = \left[\begin{array}{cc|cc} \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & \multicolumn{2}{c}{\begin{pmatrix} \underset{\sim}{F}_2'\underset{\sim}{F}_2 & \underset{\sim}{F}_2'\underset{\sim}{R}_1 \\ \underset{\sim}{R}_1'\underset{\sim}{F}_2 & \underset{\sim}{R}_1'\underset{\sim}{R}_1 \end{pmatrix}^{-1}} & \underset{\sim}{0} \\ \underset{\sim}{0} & & & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} \end{array}\right] .$$

Define

$$\begin{pmatrix} \underset{\sim}{F}_2'\underset{\sim}{F}_2 & \underset{\sim}{F}_2'\underset{\sim}{R}_1 \\ \underset{\sim}{R}_1'\underset{\sim}{F}_2 & \underset{\sim}{R}_1'\underset{\sim}{R}_1 \end{pmatrix}^{-1} \equiv \underset{\sim}{D} \equiv \begin{bmatrix} \underset{\sim}{Q}_{11} & \underset{\sim}{Q}_{12} \\ \underset{\sim}{Q}_{12} & \underset{\sim}{Q}_{22} \end{bmatrix} .$$

Then

$$\delta_A = \mathrm{tr}[\underset{\sim}{A}(\underset{\sim}{F}_2\underset{\sim}{Q}_{11}\underset{\sim}{F}_2')] .$$ Henderson <u>et</u> <u>al</u>., 1974, equ. (25)

The corresponding value of $\hat{\underset{\sim}{b}}_f$ is

$$\hat{\underset{\sim}{b}}_f = \begin{bmatrix} \hat{\underset{\sim}{b}}_{f1} \\ \hat{\underset{\sim}{b}}_{f2} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{0} \\ (\underset{\sim}{Q}_{11}\underset{\sim}{F}_2' + \underset{\sim}{Q}_{12}\underset{\sim}{R}_1')\underset{\sim}{y} \end{bmatrix} ,$$ Henderson <u>et</u> <u>al</u>., 1974, equ. (17)

i.e.,

$$\underset{\sim}{L} = \begin{bmatrix} \underset{\sim}{0} \\ \underset{\sim}{Q}_{11}\underset{\sim}{F}_2' + \underset{\sim}{Q}_{12}\underset{\sim}{R}_1' \end{bmatrix} .$$

Proof of conditions

Searle, 1968, pp. 775-777.

Proof of invariance

Henderson <u>et</u> <u>al</u>., 1974, pp. 584-586.

Importance

Method 1 cannot be used for mixed models.

But Method 1 is computationally feasible.

Method 2 <u>can</u> be used for mixed models.

And Method 2 is also computationally feasible.

Methods with a more rational basis (e.g., Maximum Likelihood) are often not computationally feasible.

4.3. Henderson's Method 3

This method uses sums of squares from the fitting constants procedure.

Utilize

$$E(\underset{\sim}{y}'\underset{\sim}{A}\underset{\sim}{y}) = tr(\underset{\sim}{A}\underset{\sim}{V}) + \underset{\sim}{\mu}'\underset{\sim}{A}\underset{\sim}{\mu} \qquad \text{(p. 9)}$$

$$= tr(\underset{\sim}{A}\underset{\sim}{V}) + tr(\underset{\sim}{A}\underset{\sim}{\mu}\underset{\sim}{\mu}')$$

$$= tr[\underset{\sim}{A}(\underset{\sim}{V} + \underset{\sim}{\mu}\underset{\sim}{\mu}')]$$

$$= tr[\underset{\sim}{A}\ E(\underset{\sim}{y}\underset{\sim}{y}')]$$

$$= tr[\underset{\sim}{A}\ E(\underset{\sim}{X}\underset{\sim}{b} + \underset{\sim}{e})(\underset{\sim}{X}\underset{\sim}{b} + \underset{\sim}{e})']$$

$$= tr\ \underset{\sim}{A}[\underset{\sim}{X}\ E(\underset{\sim}{b}\underset{\sim}{b}')\underset{\sim}{X}' + E(\underset{\sim}{e}\underset{\sim}{e}')]$$

$$= tr[\underset{\sim}{X}'\underset{\sim}{A}\underset{\sim}{X}\ E(\underset{\sim}{b}\underset{\sim}{b}')] + \sigma_e^2 tr(\underset{\sim}{A})\ . \qquad \text{[LM 444]}$$

This form is applicable to any mixed or random model.

LM 444-458 (transparencies).

4.4. Collected Results

LM Chapter 11.

Topic 5

PREDICTION and MIXED MODEL EQUATIONS

5.1. Prediction

Consider measuring intelligence in humans. Each of us has some degree of intelligence, usually called I.Q. It can never be measured exactly. As a substitute, we have test scores. Suppose we model y_{ij}, the j'th test score for the i'th person, as

$$y_{ij} = \mu + q_i + e_{ij}$$

where q_i is the person's true I.Q. and e_{ij} is a residual error term.

Mood's Exercise (Mood, 1950, p. 164, Exc. 23
Mood and Graybill, 1963, p. 195, Exc. 32)

> "23. Suppose intelligence quotients for students in a particular age group are normally distributed about a mean of 100 with standard deviation 15. The I.Q., say x_1, of a particular student is to be estimated by a test on which he scores 130. It is further given that test scores are normally distributed about the true I.Q. as a mean with standard deviation 5. What is the maximum-likelihood estimate of the student's I.Q.? (The answer is not 130.)"

This exercise raises a feature of mixed models – what is nowadays called prediction of random variables, although when 30 years ago it was referred to as estimating random variables that phrase brought forth scorn from many a mathematical statistician. Today, as prediction, it is an acceptable technique in many applications. In animal improvement plans for farm livestock, for example, it has long been known as estimating genetic (or additive genetic) merit. And quite recently (e.g., Lindley and Smith, 1972) it has attracted attention as being Bayesian estimation in linear model theory.

To solve Mood's exercise: I.Q. and score are jointly distributed with bivariate normal density:

$$\begin{bmatrix} \text{I.Q.} \\ \text{Score} \end{bmatrix} = \begin{bmatrix} q_i \\ y_{ij} \end{bmatrix} \sim N\left[\begin{pmatrix} 100 \\ 100 \end{pmatrix}, \begin{pmatrix} 15^2 & 15^2 \\ 15^2 & 15^2 + 5^2 \end{pmatrix}\right].$$

From this, the maximum likelihood estimate of the conditional mean $E(q_i|y_{ij} = 130)$ is

$$\widehat{E(q_i|y_{ij} = 130)} = 100 + \frac{15^2}{15^2 + 5^2}(130 - 100) = 127 \neq 130 .$$

This is what today is called the predicted value of q_i (given that $y_{ij} = 130$).

General formulation

$$\underset{\sim}{y} = \underbrace{\underset{\sim}{X\alpha}}_{\substack{\uparrow \\ \text{fixed} \\ \text{effects}}} + \underbrace{\underset{\sim}{Zb}}_{\substack{\uparrow \\ \text{random} \\ \text{effects}}} + \underset{\sim}{e}$$

$$E\begin{bmatrix} \underset{\sim}{b} \\ \underset{\sim}{e} \end{bmatrix} = \underset{\sim}{0} \quad \text{and} \quad \text{var}\begin{bmatrix} \underset{\sim}{b} \\ \underset{\sim}{e} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{D} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{R} \end{bmatrix}; \quad \text{and} \quad \text{var}(\underset{\sim}{y}) \equiv \underset{\sim}{V} = \underset{\sim}{ZDZ'} + \underset{\sim}{R} .$$

Then

$$\begin{bmatrix} \underset{\sim}{b} \\ \underset{\sim}{y} \end{bmatrix} \sim \left[\begin{pmatrix} \underset{\sim}{0} \\ \underset{\sim}{X\beta} \end{pmatrix}, \begin{pmatrix} \underset{\sim}{D} & \underset{\sim}{DZ'} \\ \underset{\sim}{ZD} & \underset{\sim}{ZDZ'} + \underset{\sim}{R} \end{pmatrix}\right]. \tag{1}$$

$$(\text{best predictor of } \underset{\sim}{b}) = E(\underset{\sim}{b}|\underset{\sim}{y}) . \tag{2}$$

$$\underset{\sim}{\tilde{b}} = (\text{best linear predictor of } \underset{\sim}{b}) = \underset{\sim}{DZ'}\underset{\sim}{V}^{-1}(\underset{\sim}{y} - \underset{\sim}{X\hat{\alpha}}) \tag{3}$$

with

$$\underset{\sim}{X'}\underset{\sim}{V}^{-1}\underset{\sim}{X\hat{\alpha}} = \underset{\sim}{X'}\underset{\sim}{V}^{-1}\underset{\sim}{y} . \tag{4}$$

This is true for all distributional forms of b and y. Under normality

$$\begin{pmatrix}\text{best predictor of } b \\ \text{under normality}\end{pmatrix} = \begin{pmatrix}\text{best linear} \\ \text{predictor of } b\end{pmatrix} = \tilde{b} \, . \tag{5}$$

If V is known (which it seldom is), then under normality $X\hat{\alpha}$ for $\hat{\alpha}$ of (4) is the maximum likelihood estimator of $X\alpha$ and $\tilde{b}$ is correspondingly the maximum likelihood estimator of $E(b|y)$.

Prediction of b plus linear combinations of fixed effects:

$$\text{best linear predictor (BLUP) of } w = b + K'\alpha \quad \text{is} \quad \tilde{w} = \tilde{b} + K'\hat{\alpha} \, . \tag{6}$$

Detailed derivation of (2) through (6) is available in Searle (1973, 1974).

5.2. The Mixed Model Equations

In the remainder of these notes considerable reference is made to NVC.

Aitken (GLS) equations

NVC equations

$$X'V^{-1}X\hat{\alpha} = X'V^{-1}y \, . \tag{3.1}$$

Suppose b represented fixed effects:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z \end{bmatrix} \begin{bmatrix} \alpha^o \\ b^o \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} . \tag{3.2}$$

Adding D^{-1} to this term gives Mixed Model equations

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & D^{-1} + Z'R^{-1}Z \end{bmatrix} \begin{bmatrix} \tilde{\alpha} \\ \tilde{b} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} . \tag{3.3}$$

Define as $B = B'$, the matrix on the left of (3.3). (2.61)

Alternative form considered in Harville (1977):

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{ZD} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{I} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZD} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\alpha}} \\ \tilde{\boldsymbol{\nu}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \tag{3.4}$$

with $\mathbf{D}\tilde{\boldsymbol{\nu}} = \tilde{\mathbf{b}}$.

Define as $\mathbf{C} = \mathbf{B}\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}$, the matrix on the left of (3.4). (2.62)

Singularity of $\mathbf{D}$

Equations (3.3) require $\mathbf{D}$ to be non-singular whereas (3.4) do not. By nature of its definition (1.11), $\mathbf{D}$ is customarily non-singular because, although σ_i^2 is defined for $\sigma_i^2 \geq 0$, models are usually defined only with $\sigma_i^2 > 0$. But if equations (3.3) are used in any iterative computing procedure where computed $\mathbf{D}$ is in terms of computed estimates of the σ_i^2's, any estimated σ_i^2 computed as negative or zero and accordingly given the value 0, as is often the practice, will make the computed $\mathbf{D}$ singular. This would make (3.3) unusable in a computing context, whereas it would not affect (3.4). This is the advantage of (3.4). We therefore consider both sets of equations.

Usable solutions

Table 1: NVC 52.

Generalized inverses for general solutions

$$\mathbf{C}^{-} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}^{*} \end{bmatrix} + \begin{bmatrix} \mathbf{I} \\ -\mathbf{T}^{*}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} \end{bmatrix} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}[\mathbf{I} \quad -\mathbf{X}'\mathbf{R}^{-1}\mathbf{ZDT}^{*}] \tag{2.63}$$

with $\mathbf{T}^{*} = (\mathbf{I} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZD})^{-1}$ of (2.16)

$$C^{\sim} = \begin{bmatrix} (X'R^{-1}X)^{-} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -(X'R^{-1}X)^{-}X'R^{-1}ZD \\ I \end{bmatrix} T[-Z'R^{-1}X(X'R^{-1}X)^{-} \quad I] \; . \tag{2.64}$$

with $T = (I + Z'SZD)^{-1}$ of (2.47)

and $S = R^{-1} - R^{-1}X(X'R^{-1}X)^{-}X'R^{-1}$ of (2.36)

$$B^{-} = \begin{bmatrix} I & 0 \\ 0 & D \end{bmatrix} C^{-} = \begin{bmatrix} 0 & 0 \\ 0 & DT^{*} \end{bmatrix} + \begin{bmatrix} I \\ -DT^{*}Z'R^{-1}X \end{bmatrix} (X'V^{-1}X)^{-}[I \quad -X'R^{-1}ZDT^{*}] \tag{2.65}$$

$$B^{\sim} = \begin{bmatrix} I & 0 \\ 0 & D \end{bmatrix} C^{\sim} = \begin{bmatrix} (X'R^{-1}X)^{-} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -(X'R^{-1}X)^{-}X'R^{-1}Z \\ I \end{bmatrix} DT[-Z'R^{-1}X(X'R^{-1}X)^{-} \quad I] \tag{2.67}$$

with $DT = D(I + Z'SZD)^{-1} = (D^{-1} + Z'SZ)^{-1}$.

Specific and general solutions

In general, equations $At = u$ have what can be called, for any given generalized inverse A^{-} of A

Specific solution	General Solutions
$t = A^{-}u$	$t = A^{-}u + (A^{-}A - I)z$, for arbitrary z

This duality is used (NVC 41-51) for solving (3.3) using both B^{-} and $B^{\sim}$, and for solving (3.4) using both C^{-} and $C^{\sim}$. The results, summarized in Table 2, NVC 52-53, are as one would expect: it makes no difference which generalized inverse is used, nor which equations are solved. Although this *is* to be expected, the generalized inverses look so different algebraically, that one might at first wonder if they did yield the same solutions. It is comforting to know that they do.

In essence the solutions are as follows.

For $\underset{\sim}{\alpha}$: $\quad \hat{\underset{\sim}{\alpha}} = (\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{y}$

or

an $\hat{\underset{\sim}{\alpha}} = (\underset{\sim}{X}'\underset{\sim}{R}^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{R}^{-1}\hat{\underset{\sim}{\alpha}}$

or

an $\hat{\underset{\sim}{\alpha}}^{o} = (\underset{\sim}{X}'\underset{\sim}{R}^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{R}^{-1}\underset{\sim}{X}(\hat{\underset{\sim}{\alpha}} + \underset{\sim}{w}) - \underset{\sim}{w}$

for $\underset{\sim}{w} = -\underset{\sim}{\alpha}^{o} = -\hat{\underset{\sim}{\alpha}} - [(\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X} - \underset{\sim}{I}]\underset{\sim}{z}$, with arbitrary $\underset{\sim}{z}$.

For $\underset{\sim}{b}$: $\quad \tilde{\underset{\sim}{b}} = \underset{\sim}{DZ}'\underset{\sim}{V}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}})$

For $\underset{\sim}{\nu}$ $\quad \tilde{\underset{\sim}{\nu}} = \underset{\sim}{Z}'\underset{\sim}{V}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\alpha}})$ with $\tilde{\underset{\sim}{b}} = \underset{\sim}{D}\tilde{\underset{\sim}{\nu}}$.

Uses in estimating variance components

Solution elements of the mixed model equations can be used in computational procedures for certain methods of estimating variance components. (See NVC pages 64 and 98.)

Topic 6

MAXIMUM LIKELIHOOD (ML)

6.1. Model, Likelihood and Matrices

Model: $\underset{\sim}{y} = \underbrace{\underset{\sim}{X}\underset{\sim}{\alpha}}_{\text{fixed}} + \underbrace{\underset{\sim}{Z}\underset{\sim}{b}}_{\text{random}} + \underset{\sim}{e}$

$$\sim \; N(\underset{\sim}{X\alpha}, \underset{\sim}{V}), \text{ with } \underset{\sim}{V} = \underset{\sim}{ZDZ'} + \underset{\sim}{R} .$$

Likelihood:

$$e^L = \frac{e^{-\frac{1}{2}(\underset{\sim}{y}-\underset{\sim}{X\alpha})'\underset{\sim}{V}^{-1}(\underset{\sim}{y}-\underset{\sim}{X\alpha})}}{(2\pi)^{\frac{1}{2}N}|\underset{\sim}{V}|^{\frac{1}{2}}}$$

NVC equations

$$L = -\tfrac{1}{2}N \log 2\pi - \tfrac{1}{2}\log|\underset{\sim}{V}| - \tfrac{1}{2}(\underset{\sim}{y} - \underset{\sim}{X\alpha})'\underset{\sim}{V}^{-1}(\underset{\sim}{y} - \underset{\sim}{X\alpha}) . \qquad (4.1)$$

ML (and GLS) estimator of $\underset{\sim}{\alpha}$, for known $\underset{\sim}{V}$:

$$\hat{\underset{\sim}{\alpha}} = (\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{y} .$$

Matrices:

$$\underset{\sim}{P} \equiv \underset{\sim}{V}^{-1} - \underset{\sim}{V}^{-1}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{V}^{-1} \qquad (2.26)$$

$$\underset{\sim}{Py} = \underset{\sim}{V}^{-1}(\underset{\sim}{y} - \underset{\sim}{X\hat{\alpha}}) . \qquad (3.23)$$

6.2. Maximum Likelihood Equations

Differentiating L with respect to $\underset{\sim}{\alpha}$ and each σ_i^2, $i = 0, \cdots, c$, gives the ML equations: solutions are denoted by $\tilde{\underset{\sim}{\alpha}}$ and $\tilde{\underset{\sim}{V}}$. There are various forms of the equations:

$$\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X}\tilde{\boldsymbol{\alpha}} = \mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{y} \ , \tag{4.4}$$

$$\mathrm{tr}(\tilde{\mathbf{V}}^{-1}\mathbf{Z}_i\mathbf{Z}_i') = (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\alpha}})'\tilde{\mathbf{V}}^{-1}\mathbf{Z}_i\mathbf{Z}_i'\tilde{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\alpha}}), \text{ for } i = 0, 1, \cdots, c; \tag{4.5}$$

i.e.,

$$\mathrm{tr}(\tilde{\mathbf{V}}^{-1}\mathbf{Z}_i\mathbf{Z}_i') = \mathbf{y}'\tilde{\mathbf{P}}\mathbf{Z}_i\mathbf{Z}_i'\tilde{\mathbf{P}}\mathbf{y} \ , \text{ for } i = 0, 1, \cdots, c. \tag{4.6}$$

Note that $\tilde{\boldsymbol{\alpha}}$ of (4.4) is <u>not</u> $\hat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$. Clearly, $\hat{\boldsymbol{\alpha}}$ is a function of $\mathbf{V}$, and in comparing $\tilde{\boldsymbol{\alpha}}$ with $\hat{\boldsymbol{\alpha}}$ we see that $\tilde{\boldsymbol{\alpha}}$ is $\hat{\boldsymbol{\alpha}}$ with $\mathbf{V}$ replaced by $\tilde{\mathbf{V}}$. We therefore denote $\tilde{\boldsymbol{\alpha}}$ of (4.4) as $\tilde{\hat{\boldsymbol{\alpha}}}$, and rewrite (4.4) as

$$\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X}\tilde{\hat{\boldsymbol{\alpha}}} = \mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{y} \ . \tag{4.4}$$

The left-hand side of (4.6) is

$$\begin{aligned}\mathrm{tr}(\tilde{\mathbf{V}}^{-1}\mathbf{Z}_i\mathbf{Z}_i') &\equiv \mathrm{tr}(\tilde{\mathbf{V}}^{-1}\mathbf{Z}_i\mathbf{Z}_i'\tilde{\mathbf{V}}^{-1}\tilde{\mathbf{V}}) \\ &= \mathrm{tr}(\tilde{\mathbf{V}}^{-1}\mathbf{Z}_i\mathbf{Z}_i'\tilde{\mathbf{V}}^{-1}\sum_{j=0}^{c}\tilde{\sigma}_j^2\mathbf{Z}_j\mathbf{Z}_j'), \text{ using (1.25)} \\ &= \sum_{j=0}^{c}[\mathrm{tr}(\tilde{\mathbf{V}}^{-1}\mathbf{Z}_i\mathbf{Z}_i'\tilde{\mathbf{V}}^{-1}\mathbf{Z}_j\mathbf{Z}_j')]\tilde{\sigma}_j^2 \ . \end{aligned} \tag{4.62}$$

Hence (4.6) is

$$\sum_{j=0}^{c}[\mathrm{tr}(\tilde{\mathbf{V}}^{-1}\mathbf{Z}_i\mathbf{Z}_i'\tilde{\mathbf{V}}^{-1}\mathbf{Z}_j\mathbf{Z}_j']\tilde{\sigma}_j^2 = \mathbf{y}'\tilde{\mathbf{P}}\mathbf{Z}_i\mathbf{Z}_i'\tilde{\mathbf{P}}\mathbf{y} \ , \quad \text{for } i,j = 0, \cdots, c.$$

These equations can be written in vector form as

$$\{\mathrm{tr}(\tilde{\mathbf{V}}^{-1}\mathbf{Z}_i\mathbf{Z}_i'\tilde{\mathbf{V}}^{-1}\mathbf{Z}_j\mathbf{Z}_j')\}\tilde{\boldsymbol{\sigma}}^2 = \{\mathbf{y}'\tilde{\mathbf{P}}\mathbf{Z}_i\mathbf{Z}_i'\tilde{\mathbf{P}}\mathbf{y}\} \ , \quad \text{for } i,j = 0, \cdots, c. \tag{4.63}$$

This form of the equations readily provides an iterative scheme for obtaining a solution. With an initial value for $\mathring{\boldsymbol{\sigma}}^2$, calculate $\mathbf{V}$ and the matrix on the left and also calculate $\mathbf{P}$ and the vector on the right. Solve for $\tilde{\boldsymbol{\sigma}}^2$ and repeat. When convergence is reached, use the final value of $\tilde{\boldsymbol{\sigma}}^2$ to compute a final $\tilde{\mathbf{V}}$ and thence

$$\tilde{\hat{\boldsymbol{\alpha}}} = (\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{y} \ .$$

6.3. A Computational Note

For any matrix $\underset{\sim}{A}$,

$$\begin{aligned}\operatorname{tr}(\underset{\sim}{AA}') &= \Sigma(\text{diagonal elements of } \underset{\sim}{AA}')\\ &= \Sigma(\text{inner product of each row of } \underset{\sim}{A} \text{ with itself})\\ &= \Sigma\Sigma a_{ij}^2\\ &= \text{sum of squares of elements of } \underset{\sim}{A}.\end{aligned}$$

Write this as

$$\operatorname{tr}(\underset{\sim}{AA}') = \operatorname{ssqe}(\underset{\sim}{A}) .$$

Then, because in (4.63)

$$\begin{aligned}\operatorname{tr}(\underset{\sim}{V}^{-1}\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\underset{\sim}{V}^{-1}\underset{\sim}{Z}_j\underset{\sim}{Z}_j') &= \operatorname{tr}(\underset{\sim}{Z}_i'\underset{\sim}{V}^{-1}\underset{\sim}{Z}_j\underset{\sim}{Z}_j'\underset{\sim}{V}^{-1}\underset{\sim}{Z}_i)\\ &= \operatorname{tr}[\underset{\sim}{Z}_i'\underset{\sim}{V}^{-1}\underset{\sim}{Z}_j(\underset{\sim}{Z}_i'\underset{\sim}{V}^{-1}\underset{\sim}{Z}_j)']\end{aligned}$$

equations (4.63) can be written as

$$\{\operatorname{ssqe}(\underset{\sim}{Z}_i'\tilde{\underset{\sim}{V}}^{-1}\underset{\sim}{Z}_j)\}\tilde{\underset{\sim}{\sigma}} = \{\operatorname{ssqe}(\underset{\sim}{Z}_i'\underset{\sim}{Py})\} \quad \text{for } i,j = 0, \cdots, c.$$

This computing principle is utilized, for example, in SAS VARCOMP.

6.4. Ratios of Variance Components

Hartley and Rao (1967) is the original paper for ML. It deals with estimating

$$\underset{\sim}{\alpha},\ \sigma_0^2 \text{ and } \underset{\sim}{\gamma} = \{\gamma_i\} = \{\sigma_i^2/\sigma_0^2\},\ \text{for } i = 1, \cdots, c.$$

Their equations, using

$$\underset{\sim}{H} = \underset{\sim}{V}/\sigma_0^2 \tag{1.21}$$

are

$$\underset{\sim}{X}'\underset{\sim}{\tilde{H}}^{-1}\underset{\sim}{X}\underset{\sim}{\tilde{\hat{\alpha}}} = \underset{\sim}{X}'\underset{\sim}{\tilde{H}}^{-1}\underset{\sim}{y} \tag{4.11}$$

$$\tilde{\sigma}_0^2 = (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\tilde{\hat{\alpha}}})'\underset{\sim}{\tilde{H}}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\tilde{\hat{\alpha}}})/N \tag{4.12}$$

$$\mathrm{tr}(\underset{\sim}{\tilde{H}}^{-1}\underset{\sim}{Z}_i\underset{\sim}{Z}_i') = (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\tilde{\hat{\alpha}}})'\underset{\sim}{\tilde{H}}^{-1}\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\underset{\sim}{\tilde{H}}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\tilde{\hat{\alpha}}})/\tilde{\sigma}_0^2 \quad \text{for } i = 1, \cdots, c. \tag{4.13}$$

Equivalence of (4.4) and (4.5) to (4.11) - (4.13) is shown in NVC 60-61. In particular, (4.12) and (4.13) are equivalent to

$$\mathrm{tr}(\underset{\sim}{\tilde{H}}^{-1}\underset{\sim}{Z}_i\underset{\sim}{Z}_i') = (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\tilde{\hat{\alpha}}})'\underset{\sim}{\tilde{H}}^{-1}\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\underset{\sim}{\tilde{H}}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\tilde{\hat{\alpha}}})/\tilde{\sigma}_0^2 \quad \text{for } i = 0, 1, \cdots, c, \tag{4.16}$$

which is also equivalent to (4.5). Furthermore, on defining

$$\underset{\sim}{Q} = \underset{\sim}{H}^{-1} - \underset{\sim}{H}^{-1}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{H}^{-1}\underset{\sim}{X})^-\underset{\sim}{X}'\underset{\sim}{H}^{-1} = \underset{\sim}{P}\sigma_0^2 ,$$

(4.63) becomes

$$\{\mathrm{tr}(\underset{\sim}{\tilde{H}}^{-1}\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\underset{\sim}{\tilde{H}}^{-1}\underset{\sim}{Z}_j\underset{\sim}{Z}_j')\}\underset{\sim}{\tilde{\sigma}}^2 = \{\underset{\sim}{y}'\underset{\sim}{\tilde{Q}}\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\underset{\sim}{\tilde{Q}}\underset{\sim}{y}\} \quad \text{for } i = 0, 1, \cdots, c.$$

This may have some computational advantage in that

$$\underset{\sim}{H} = \sum_{i=1}^{c} \gamma_i\underset{\sim}{Z}_i\underset{\sim}{Z}_i' + \underset{\sim}{I}_N \tag{1.22}$$

might be numerically more stable for inversion than $\underset{\sim}{V}$.

6.5. Using the Mixed Model Equations

Equations (4.12) and (4.13) can be expressed in terms of $\underset{\sim}{\hat{\alpha}}$ and $\underset{\sim}{\tilde{b}}$ derived from the Mixed Model Equations, to yield an iterative process as shown on NVC 64. It always gives positive estimates for the variance components — but its convergence properties are unknown.

6.6. ML Estimators

ML estimators of σ^2's are to be distinguished from solutions of ML equations. Solutions might be negative, but the ML procedure demands that ML estimators maximize the likelihood over the parameter space which, for $\dot{\sigma}^2$ is $\sigma_e^2 > 0$ and $\sigma_i^2 \geq 0$ for $i = 1, \cdots, c$. ML estimators must satisfy these same conditions, $\tilde{\sigma}_0^2 > 0$ and $\tilde{\sigma}_i^2 \geq 0$ for $i = 1, \cdots, c$.

These conditions are a difficulty that must be taken into account in computer programs that are used for solving the ML equations to obtain ML estimators. Customarily, any $\tilde{\sigma}_i^2$ that is computed as a negative value is put equal to zero — an action which has the effect, of course, of altering the model being used. It also raises the further difficulty of having a computer program which, for any $\tilde{\sigma}_i^2$ that has been put equal to zero after some iteration, enables that $\tilde{\sigma}_i^2$ to come back into the calculations again at some later iteration if it were then to be positive. Computing difficulties of this nature are considered in such papers as Hemmerle and Hartley (1973) and Jennrich and Sampson (1976).

6.7. Information Matrices

Using the likelihood function (4.1), Searle (1970) has shown that the information matrix for $\dot{\sigma}^2$ of (1.27) is

$$I\begin{pmatrix} \alpha \\ \dot{\sigma}^2 \end{pmatrix} = \frac{1}{2}\begin{bmatrix} X'V^{-1}X & 0 \\ 0 & I(\dot{\sigma}^2) \end{bmatrix}$$

where

$$I(\dot{\sigma}^2) = \frac{1}{2}\left\{ \operatorname{tr}\left(V^{-1}\frac{\partial V}{\partial \sigma_i^2} V^{-1}\frac{\partial V}{\partial \sigma_j^2}\right)\right\} \tag{4.32}$$

$$= \frac{1}{2}\left\{ \operatorname{tr}\left(V^{-1}Z_iZ_i'V^{-1}Z_jZ_j'\right)\right\}, \quad \text{for } i,j = 0, 1, \cdots, c. \tag{4.33}$$

The large sample variances are obtained from this by inversion:

$$\operatorname{var}(\tilde{\sigma}^2) = [I(\dot{\sigma}^2)]^{-1} .$$

The lower half of Table 4.1 on NVC 56 shows the equation numbers in NVC where various facets of information matrices are developed in detail (pages 66 through 79 – the algebra is extensive).

6.8. Computing Algorithms

The purpose of NVC is to present algebraic derivations, not computer techniques. Nevertheless, Newton-Rhapson and Fisher scoring are discussed briefly at NVC 80-85.

Topic 7

RESTRICTED MAXIMUM LIKELIHOOD (REML)

7.1. Error Contrasts and Their Likelihood

Thompson (1962): maximize that portion of the likelihood which is invariant to the mean.

Patterson and Thompson (1971): generalized.

Corbeil and Searle (1976): specific algorithm.

Error contrast: Harville (1974)

$$E(\underset{\sim}{k}'\underset{\sim}{y}) = \underset{\sim}{k}'\underset{\sim}{X}\underset{\sim}{\alpha}$$

$\underset{\sim}{k}'\underset{\sim}{y}$ is an error contrast when

$$E(\underset{\sim}{k}'\underset{\sim}{y}) = \underset{\sim}{0} \text{ for all } \underset{\sim}{\alpha} ;$$

$$\Rightarrow \underset{\sim}{k}'\underset{\sim}{X} = \underset{\sim}{0} .$$

$\underset{\sim}{X}$ is $N \times p$ with rank p^* .

Therefore there are only $N - p^*$ LIN k's for $\underset{\sim}{k}'\underset{\sim}{X} = \underset{\sim}{0}$.

Hence we consider

$$\underset{\sim}{K}'\underset{\sim}{X} = \underset{\sim}{0}, \quad \text{with} \quad r(\underset{\sim}{K}') = N - p^*, \text{ of full row rank.}$$

Likelihood

The likelihood of $\underset{\sim}{K}'\underset{\sim}{y}$ is

$$L(\underset{\sim}{K}'\underset{\sim}{y}) = -\tfrac{1}{2}(N - p^*)\log 2\pi - \tfrac{1}{2}\log|\underset{\sim}{K}'\underset{\sim}{VK}| - \tfrac{1}{2}\underset{\sim}{y}'\underset{\sim}{K}(\underset{\sim}{K}'\underset{\sim}{VK})^{-1}\underset{\sim}{K}'\underset{\sim}{y} .$$

It has many equivalent forms (NVC 88-91). The most convenient is

$$L(\underset{\sim}{K}'\underset{\sim}{y}) = L_1 + \text{constant not dependent on } \underset{\sim}{\alpha} \text{ or } \underset{\sim}{\sigma}^2$$

for

$$L_1 = -\tfrac{1}{2}\log|\underset{\sim}{V}| - \tfrac{1}{2}\log|\underset{\sim}{X}^{*\prime}\underset{\sim}{V}^{-1}\underset{\sim}{X}^*| - \tfrac{1}{2}\underset{\sim}{y}'\underset{\sim}{Py} \qquad (5.2)$$

where $\mathbf{X}^*$ is any $p^* = r(\mathbf{X})$ LIN columns of $\mathbf{X}$. Derivation of this result involves using

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' \tag{2.17}$$

$$\mathbf{A}, \text{ such that } \mathbf{M} = \mathbf{AA}' \quad \text{and} \quad \mathbf{A}'\mathbf{A} = \mathbf{I}_{N-p^*} \,. \tag{2.22}$$

$\mathbf{K}'\mathbf{X} = \mathbf{0}$ if and only if $\mathbf{K}' = \mathbf{K}'\mathbf{M}$

[This alters $\mathbf{W}'$ of NVC (2.73) to be $\mathbf{K}'$]

$\mathbf{K}' = \mathbf{K}'\mathbf{AA}'$ with $\mathbf{K}'\mathbf{A}$ nonsingular [Lemma 2.3, NVC 25]

$$\mathbf{P} = \mathbf{A}(\mathbf{A}'\mathbf{VA})^{-1}\mathbf{A}' \tag{2.35}$$

$$\mathbf{K}(\mathbf{K}'\mathbf{VK})^{-1}\mathbf{K}' = \mathbf{P} \,. \tag{2.74}$$

7.2. The REML Equations

Differentiating (5.2) with respect to σ_i^2, using

$$\frac{\partial}{\partial\sigma_i^2}\log|\mathbf{V}| = \mathrm{tr}\left(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\sigma_i^2}\right) = \mathrm{tr}(\mathbf{V}^{-1}\mathbf{Z}_i\mathbf{Z}_i')\,, \tag{2.4}$$

$$\partial\mathbf{P} = -\mathbf{A}(\mathbf{A}'\mathbf{VA})^{-1}\mathbf{A}'(\partial\mathbf{V})\mathbf{A}(\mathbf{A}'\mathbf{VA})^{-1}\mathbf{A}' = -\mathbf{P}(\partial\mathbf{V})\mathbf{P}\,, \tag{2.80}$$

and

$$\mathbf{P}^* = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{V}^{-1}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{V}^{-1} = \mathbf{P}\,,$$

leads to the REML equations

$$\mathrm{tr}(\hat{\mathbf{P}}\mathbf{Z}_i\mathbf{Z}_i') = \{\mathbf{y}'\hat{\mathbf{P}}\mathbf{Z}_i\mathbf{Z}_i'\hat{\mathbf{P}}\mathbf{y}\} \quad \text{for } i = 0, 1, \cdots, c. \tag{5.17}$$

Since $\mathbf{P} = \mathbf{PVP}$,

$$\begin{aligned}\mathrm{tr}(\mathbf{PZ}_i\mathbf{Z}_i') &= \mathrm{tr}(\mathbf{PVPZ}_i\mathbf{Z}_i') \\ &= \mathrm{tr}(\mathbf{PZ}_i\mathbf{Z}_i'\mathbf{PV}) \\ &= \mathrm{tr}(\mathbf{PZ}_i\mathbf{Z}_i'\mathbf{P}\sum_{j=0}^{c}\sigma_j^2\mathbf{Z}_j\mathbf{Z}_j') \\ &= \sum_{j=0}^{c}\mathrm{tr}(\mathbf{PZ}_i\mathbf{Z}_i'\mathbf{PZ}_j\mathbf{Z}_j')\sigma_j^2\end{aligned}$$

so enabling (5.17) to be expressed as

$$\left\{\mathrm{tr}(\hat{\underset{\sim}{P}}\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\hat{\underset{\sim}{P}}\underset{\sim}{Z}_j\underset{\sim}{Z}_j')\right\}\hat{\underset{\sim}{\sigma}}^2 = \left\{\underset{\sim}{y}'\hat{\underset{\sim}{P}}\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\hat{\underset{\sim}{P}}\underset{\sim}{y}\right\}, \quad \text{for } i,j = 0, 1, \cdots, c. \tag{5.19}$$

These can also be expressed as

$$\{\mathrm{ssqe}(\underset{\sim}{Z}_i'\hat{\underset{\sim}{P}}\underset{\sim}{Z}_j)\} = \{\mathrm{ssqe}(\underset{\sim}{Z}_i'\hat{\underset{\sim}{P}}\underset{\sim}{y})\} \quad \text{for } i,j = 0, 1, \cdots, c.$$

Direct derivation from ML:

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\alpha} + \underset{\sim}{Z}\underset{\sim}{b} + \underset{\sim}{e} \sim \mathcal{N}(\underset{\sim}{X}\underset{\sim}{\alpha}, \underset{\sim}{V}), \tag{5.26}$$

with ML equations

$$\mathrm{tr}(\tilde{\underset{\sim}{V}}^{-1}\underset{\sim}{Z}_i\underset{\sim}{Z}_i') = \underset{\sim}{y}'\tilde{\underset{\sim}{P}}\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\tilde{\underset{\sim}{P}}\underset{\sim}{y}, \quad \text{for } i = 0, \cdots, c. \tag{5.25}$$

REML is just ML applied to

$$\underset{\sim}{K}'\underset{\sim}{y} = \underset{\sim}{K}'\underset{\sim}{Z}\underset{\sim}{b} + \underset{\sim}{K}'\underset{\sim}{e} \sim \mathcal{N}(\underset{\sim}{0}, \underset{\sim}{K}'\underset{\sim}{V}\underset{\sim}{K}). \tag{5.27}$$

Compare (5.26) and (5.27): in (5.25)

$$\begin{array}{lccl} \text{change} & \underset{\sim}{y} & \text{to} & \underset{\sim}{K}'\underset{\sim}{y} \\ & \underset{\sim}{X} & \text{to} & \underset{\sim}{K}'\underset{\sim}{X} = \underset{\sim}{0} \\ & \underset{\sim}{Z} & \text{to} & \underset{\sim}{K}'\underset{\sim}{Z} \\ & \underset{\sim}{V} & \text{to} & \underset{\sim}{K}'\underset{\sim}{V}\underset{\sim}{K} \\ & \underset{\sim}{P} & \text{to} & (\underset{\sim}{K}'\underset{\sim}{V}\underset{\sim}{K})^{-1} \end{array}$$

and so the REML equations are, for $i = 0, 1, \cdots, c$

$$\mathrm{tr}[(\underset{\sim}{K}'\hat{\underset{\sim}{V}}\underset{\sim}{K})^{-1}\underset{\sim}{K}'\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\underset{\sim}{K}] = \underset{\sim}{y}'\underset{\sim}{K}(\underset{\sim}{K}'\hat{\underset{\sim}{V}}\underset{\sim}{K})^{-1}\underset{\sim}{K}'\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\underset{\sim}{K}(\underset{\sim}{K}'\hat{\underset{\sim}{V}}\underset{\sim}{K})^{-1}\underset{\sim}{K}'\underset{\sim}{y},$$

i.e.,

$$\mathrm{tr}[\underset{\sim}{K}(\underset{\sim}{K}'\hat{\underset{\sim}{V}}\underset{\sim}{K})^{-1}\underset{\sim}{K}'\underset{\sim}{Z}_i\underset{\sim}{Z}_i'] = \underset{\sim}{y}'\hat{\underset{\sim}{P}}\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\hat{\underset{\sim}{P}}\underset{\sim}{y},$$

or

$$\mathrm{tr}(\hat{\underset{\sim}{P}}\underset{\sim}{Z}_i\underset{\sim}{Z}_i') = \underset{\sim}{y}'\hat{\underset{\sim}{P}}\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\hat{\underset{\sim}{P}}\underset{\sim}{y},$$

as in (5.17); from which (5.19) is available.

7.3. Other Topics

Single equation for $\hat{\sigma}_0^2$

$$\hat{\sigma}_0^2 = \frac{(\underset{\sim}{y} - \underset{\sim}{X}\hat{\hat{\underset{\sim}{\alpha}}})'\underset{\sim}{H}^{-1}(\underset{\sim}{y} - \underset{\sim}{X}\hat{\hat{\underset{\sim}{\alpha}}})}{N - p^*} . \tag{5.22}$$

Equations for ratios: NVC 94-95.

Comparisons with ML: NVC 96.

Using mixed model equations: NVC 96-98, and 103-104.

Information matrices: NVC 99-105.

Table on NVC 87 summarizes.

Computing algorithms: NVC 106-107.

Corbeil and Searle's REML: NVC 108-109.

Topic 8

MINQUE and MIVQUE

8.1. MINQUE

Minimum norm, quadratic, unbiased, estimation.

Quadratic estimation

For any given $\mathbf{p}'$, estimate $\mathbf{p}'\dot{\boldsymbol{\sigma}}^2$.

$$\widehat{\mathbf{p}'\dot{\boldsymbol{\sigma}}^2} = \mathbf{y}'\mathbf{A}\mathbf{y}, \quad \text{with} \quad \mathbf{A} = \mathbf{A}' .$$

Choose $\mathbf{A}$ to satisfy certain criteria.

Invariance to fixed effects

$$\mathbf{A}\mathbf{X} = \mathbf{0} \qquad \Rightarrow \mathbf{A} = \mathbf{A}' = \mathbf{A}\mathbf{M} = \mathbf{M}\mathbf{A} = \mathbf{M}\mathbf{A}\mathbf{M} .$$

Unbiasedness

$$\widehat{\mathbf{p}'\dot{\boldsymbol{\sigma}}^2} = \mathbf{y}'\mathbf{A}\mathbf{y}$$

$$\mathbf{p}'\dot{\boldsymbol{\sigma}}^2 = E(\mathbf{y}'\mathbf{A}\mathbf{y}) = \mathrm{tr}(\mathbf{A}\mathbf{V}) = \mathrm{tr}(\mathbf{A}\Sigma\sigma_i^2\mathbf{Z}_i\mathbf{Z}_i')$$

$$\Sigma p_i\sigma_i^2 = \Sigma\mathrm{tr}(\mathbf{A}\mathbf{Z}_i\mathbf{Z}_i')\sigma_i^2$$

$$\Rightarrow \quad p_i = \mathrm{tr}(\mathbf{A}\mathbf{Z}_i\mathbf{Z}_i') .$$

Minimum norm

If the $\mathbf{b}_i$'s in $\mathbf{b}$ were known, with $E(\mathbf{b}_i) = 0$ and $\mathrm{var}(\mathbf{b}_i) = \sigma_i^2\mathbf{I}_{q_i}$, Rao suggests that a "natural" estimator would be

$$\widetilde{\mathbf{p}'\dot{\boldsymbol{\sigma}}^2} = \sum_{i=0}^{c} \frac{p_i(\mathbf{b}_i'\mathbf{b}_i)}{q_i} = \dot{\mathbf{b}}'\dot{\boldsymbol{\Delta}}\dot{\mathbf{b}} \tag{6.3}$$

for

$$\dot{\boldsymbol{\Delta}} = \mathrm{diag}\left\{\frac{p_i}{q_i}\mathbf{I}_{q_i}\right\} \quad \text{for } i = 0, 1, \cdots, c. \tag{6.5}$$

But

$$\widehat{\underset{\sim}{p}'\dot{\underset{\sim}{\sigma}}^2} = \underset{\sim}{y}'\underset{\sim}{A}\underset{\sim}{y} = \dot{\underset{\sim}{b}}'\dot{\underset{\sim}{Z}}'\underset{\sim}{A}\dot{\underset{\sim}{Z}}\dot{\underset{\sim}{b}} \tag{6.6}$$

and

$$\widehat{\underset{\sim}{p}'\dot{\underset{\sim}{\sigma}}^2} - \widetilde{\underset{\sim}{p}'\dot{\underset{\sim}{\sigma}}^2} = \dot{\underset{\sim}{b}}'(\dot{\underset{\sim}{Z}}'\underset{\sim}{A}\dot{\underset{\sim}{Z}} - \dot{\underset{\sim}{\Delta}})\dot{\underset{\sim}{b}} \ . \tag{6.7}$$

A weighted Euclidean norm of the matrix in (6.7) is used with weights $w_0, w_1, \cdots, w_c$, corresponding to (a priori values of) $\sigma_0^2, \sigma_1^2, \cdots, \sigma_c^2$. Define

$$\dot{\underset{\sim}{D}}_w = \operatorname{diag}\{w_i \underset{\sim}{I}_{q_i}\} \quad \text{for } i = 0, 1, \cdots, c$$

and

$$\underset{\sim}{V}_w = \dot{\underset{\sim}{Z}}\dot{\underset{\sim}{D}}_w\dot{\underset{\sim}{Z}} \ . \tag{6.8}$$

The norm that is minimized is

$$\|\dot{\underset{\sim}{D}}_w^{\frac{1}{2}}(\dot{\underset{\sim}{Z}}'\underset{\sim}{A}\dot{\underset{\sim}{Z}} - \dot{\underset{\sim}{\Delta}})\dot{\underset{\sim}{D}}_w^{\frac{1}{2}}\| = \operatorname{tr}[\dot{\underset{\sim}{D}}_w^{\frac{1}{2}}(\dot{\underset{\sim}{Z}}'\underset{\sim}{A}\dot{\underset{\sim}{Z}} - \dot{\underset{\sim}{\Delta}})\dot{\underset{\sim}{D}}_w^{\frac{1}{2}}]^2 \ .$$

Estimation equations

After considerable algebra, the equations for estimating $\dot{\underset{\sim}{\sigma}}^2$ turn out to be

$$\{\operatorname{tr}(\underset{\sim}{P}_w\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\underset{\sim}{P}_w\underset{\sim}{Z}_j\underset{\sim}{Z}_j')\}\hat{\dot{\underset{\sim}{\sigma}}}^2 = \{\underset{\sim}{y}'\underset{\sim}{P}_w\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\underset{\sim}{P}_w\underset{\sim}{y}\} \tag{6.26}$$

for $i,j = 0, 1, \cdots, c$. These are the MINQUE estimators of the variance components, using weights w_i in the norm. Since these weights are pre-assigned numbers, $\underset{\sim}{V}_w$ and hence $\underset{\sim}{P}_w$ are matrices that can be calculated, and so the solutions to (6.26) can also be calculated — provided the w_i's are such that the matrix on the left side of (6.26) is nonsingular.

Note: MINQUE does not involve iteration.
MINQUE depends on the pre-assigned weights, corresponding to the σ^2's.
MINQUE involves no distributional assumptions.

8.2. MINQUE0

Using zero for all w_i's save w_0, i.e., $w_i = 0$ for $i = 1, \cdots, c$ is of interest because this gives $\underset{\sim}{V}_w = w_0\underset{\sim}{I}$ and $\underset{\sim}{P}_w = \underset{\sim}{M}/w_0$. Then (6.26) is

$$\hat{\underset{\sim}{\sigma}}^2 = \{\mathrm{tr}(\underset{\sim}{M}\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\underset{\sim}{M}\underset{\sim}{Z}_j\underset{\sim}{Z}_j')\}^{-1}\{\underset{\sim}{y}'\underset{\sim}{M}\underset{\sim}{Z}_i\underset{\sim}{Z}_i'\underset{\sim}{M}\underset{\sim}{y}\} \ . \tag{6.27}$$

These are the estimators suggested by Rao (1970) in the first of his four papers on MINQUE.

MINQUE0 has been called MIVQUE0 by Goodnight (1978, 1979), in keeping with MINQUE being MIVQUE under normality. Without using the name MINQUE0, Seely (1971, equation (6)) has it as a method of estimation, Corbeil and Searle (1976a) have it as the starting point of the (iterative) REML procedure, and Hartley et al. (1978, equation (10)) espouse its use on grounds of relatively easy computability, a feature that is promoted by Goodnight (1978). Reconciliation with (6.27) of the Corbeil and Searle (1976a) and the Hartley et al. (1978) descriptions is given at NVC 119-122.

8.3. I-MINQUE (Iterative MINQUE)

Equations (6.26) for MINQUE are the same as (5.19) for REML except that $\underset{\sim}{P}_w$ is used in the MINQUE equations in place of $\hat{\underset{\sim}{P}}$ in the REML equations — and the MINQUE equations have a direct solution whereas the REML equations are solved iteratively. And in $\underset{\sim}{P}_w$, $\underset{\sim}{w}$ corresponds to an à priori value of $\underset{\sim}{\sigma}^2$. But if $\underset{\sim}{w}$ is thought of as an initial value for $\underset{\sim}{\sigma}^2$ in the MINQUE equations, from which those equations are solved iteratively, then that procedure is called I-MINQUE; and, clearly, it is the same as REML. Furthermore, when using $\underset{\sim}{w}$ in MINQUE the MINQUE is equivalent to a first iterate of REML starting with $\underset{\sim}{w}$. Therefore

$$\text{I-MINQUE} = \text{REML} \tag{6.29}$$

$$\text{a MINQUE} = \text{a first iterate of REML.} \tag{6.30}$$

REML, of course, is based on normality; MINQUE requires no such assumption. But Brown (1978) shows that MINQUE and I-MINQUE estimators are asymptotically normal.

8.4. MIVQUE

On assuming normality, minimum variance, quadratic unbiased estimators are the same as MINQUE estimators; and so

$$\text{MIVQUE (under normality)} = \text{MINQUE} .$$

Topic 9

DISPERSION MEAN MODEL

9.1. The Model

The customary general linear model is $\underset{\sim}{y} \sim (\underset{\sim}{Xb}, \underset{\sim}{V})$. The variance component problem can also be written as a linear model

$$\underset{\sim}{\mathcal{Y}} \sim (\underset{\sim}{\mathcal{X}\dot{\sigma}^2}, \underset{\sim}{W}) \ , \tag{7.2}$$

where $\underset{\sim}{\mathcal{Y}}$, $\underset{\sim}{\mathcal{X}}$ and $\underset{\sim}{W}$ involve a variety of matrix operations.

9.2. Estimation from the Model

OLS yields MINQUE0 (7.15)

GLS yields MINQUE .

9.3. Derivation of the Model NVC 123-126

The following three matrix operators are used.

Kronecker product: $\underset{\sim}{A} \otimes \underset{\sim}{B} = \{a_{ij}\underset{\sim}{B}\}$

Direct sum: $\underset{\sim}{A} \oplus \underset{\sim}{B} = \begin{bmatrix} \underset{\sim}{A} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{B} \end{bmatrix}$

vec: $\text{vec}\begin{bmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 3 \\ 6 \\ 4 \\ 7 \end{bmatrix}$

and also the vec-permutation matrix of order N^2:

$$\underset{\sim}{S}_N = \underset{\sim}{I}_{N,N} \quad \text{e.g., } \underset{\sim}{I}_{(3,3)} = \left[\begin{array}{ccc|ccc|ccc} 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \hline \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \hline \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{array}\right].$$

Then

$$\underset{\sim}{A}\underset{\sim}{A}' = \underset{\sim}{M} \quad \text{and} \quad \underset{\sim}{A}'\underset{\sim}{A} = \underset{\sim}{I} \tag{7.7}$$

$$\underset{\sim}{\mathcal{Y}} = \underset{\sim}{A}'\underset{\sim}{y} \otimes \underset{\sim}{A}'\underset{\sim}{y} \tag{7.9}$$

$$\underset{\sim}{\psi} = [\text{vec}(\underset{\sim}{Z}_0\underset{\sim}{Z}_0') \cdots \text{vec}(\underset{\sim}{Z}_c\underset{\sim}{Z}_c')]$$

$$\underset{\sim}{\chi} = (\underset{\sim}{A}' \otimes \underset{\sim}{A}')\underset{\sim}{\psi}$$

$$\gamma_i = \text{kurtosis of population from which } \underset{\sim}{b}_i \text{ comes}$$

$$\underset{\sim}{\Gamma} = \bigoplus_{i=0}^{c} \gamma_i \sigma_i^4 \underset{\sim}{I}_{q_i} = \text{diag}\left\{\gamma_i \sigma_i^4 \underset{\sim}{I}_{q_i}\right\} \tag{7.18}$$

$$\underset{\sim}{F} = (\underset{\sim}{V} \otimes \underset{\sim}{V})(\underset{\sim}{I} + \underset{\sim}{S}_N) + (\underset{\sim}{Z} \otimes \underset{\sim}{Z})\underset{\sim}{\Gamma}(\underset{\sim}{Z}' \otimes \underset{\sim}{Z}') \tag{7.17}$$

$$\underset{\sim}{W} = (\underset{\sim}{A}' \otimes \underset{\sim}{A}')\underset{\sim}{F}(\underset{\sim}{A} \otimes \underset{\sim}{A}) \tag{7.16}$$

<u>9.4. A Modified Model</u> NVC 132-136

$$\underset{\sim}{\mathcal{Y}}_1 = (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\alpha}) \otimes (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\alpha}) \tag{7.37}$$

$$\sim (\underset{\sim}{\psi}\underset{\sim}{\dot\sigma}^2, \underset{\sim}{F}) .$$

Assume $\underset{\sim}{\alpha}$ is known; carry out GLS on this model and then replace $\underset{\sim}{X}\underset{\sim}{\alpha}$ by $\underset{\sim}{X}\hat{\underset{\sim}{\alpha}} = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{y}$.

This gives ML.